Estimating AFQT Scores for National Educational Longitudinal Study (NELS) Respondents

M. Rebecca Kilburn

Lawrence M. Hanser

Jacob A. Klerman

National Defense Research Institute • Arroyo Center

50th Year RAND

A NATIONAL RESOURCE

1948–1998

The research described in this report was conducted within two RAND federally funded research and development centers: The National Defense Research Institute, sponsored by the Office of the Secretary of Defense, the Joint Staff, the unified commands, and the defense agencies, Contract DASW01-95-C-0059; and by the Arroyo Center sponsored by the United States Army, Contract DASW01-96-C-0004.

Library of Congress Cataloging-in-Publication Data

Kilburn, M. Rebecca.
Estimating AFQT scores for national educational longitudinal study (NELS) respondents / M. Rebecca Kilburn, Lawrence M. Hanser, and Jacob A. Klerman.
p. cm.
"Prepared for the Office of the Secretary of Defense and for the U.S. Army by RAND's National Defense Research Institute and by RAND's Arroyo Center."
"MR-818-OSD/A."
Includes bibliographical references.
ISBN 0-8330-2581-3 (alk. paper)
1. United States—Armed Forces—Recruiting, enlistment, etc.—Forecasting.
I. Hanser, Lawrence M. II. Klerman, Jacob Alex. III. United States. Dept. of Defense. Office of the Secretary of Defense. IV. RAND Corporation.
V. National Defense Research Institute (U.S.). VI. Arroyo Center. VII. Title.
UB323.K527 1998
355.2'234'0973—dc21 97-52245
CIP

RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.

Published 1998 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1333 H St., N.W., Washington, D.C. 20005-4707
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact Distribution Services: Telephone: (310) 451-7002; Fax: (310) 451-6915; Internet: order@rand.org

Estimating AFQT Scores for National Educational Longitudinal Study (NELS) Respondents

M. Rebecca Kilburn
Lawrence M. Hanser
Jacob A. Klerman

MR-818-OSD/A

Prepared for the
Office of the Secretary of Defense
United States Army

National Defense Research Institute • Arroyo Center

RAND

PREFACE

This document describes a methodology for imputing AFQT scores for members of the National Educational Longitudinal Study (NELS) sample. It then uses that methodology to explore some implications of test score trends for military recruiting. A forthcoming report[1] uses these estimated test scores in models of individual enlistment behavior. The report will be of interest both to individuals concerned with the methodological issues related to imputing test scores and to those setting recruiting policy.

This research, part of the "Recruiting Policy and Resources" project, was conducted for the Under Secretary of Defense for Personnel and Readiness and for the Deputy Chief of Staff for Personnel, U.S. Army. The project was executed jointly by the Forces and Resources Policy Center of RAND's National Defense Research Institute (NDRI) and by the Manpower and Training Program of RAND's Arroyo Center. NDRI and the Arroyo Center are both federally funded research and development centers, the first sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Commands, and the defense agencies, and the second sponsored by the United States Army.

[1]M. Rebecca Kilburn and Jacob A. Klerman, *Enlistment Decisions in the 1990s: Evidence from Individual-Level Data*, Santa Monica, CA: RAND, MR-944-OSD/A, forthcoming.

CONTENTS

FIGURES

TABLES

SUMMARY

In studies using the 1980 wave of the National Longitudinal Survey of Youth (NLSY), Hosek and Peterson (1985, 1990) showed that although eligibility for enlistment is conditional on surpassing a minimum score on the Armed Forces Qualification Test (AFQT), the probability that individuals enlist falls as AFQT score rises. In addition, they found the decision to enlist to be related to a number of other individual characteristics. The NLSY sample was representative of a cohort of youths in 1980. Since that time, changes have taken place that would be expected to influence individual enlistment probabilities. Therefore, DoD needs to reestimate these models using more current data.

This paper reports on the first segment of a project that estimates the determinants of individual enlistment decisions using the 1992 second follow-up of the National Educational Longitudinal Study (NELS). The NELS sample contains youths who were high school seniors in 1992. The project's first segment estimates AFQT scores for NELS respondents so that we can replicate the Hosek and Peterson studies with NELS, and it uses the results of this estimation to draw some implications for recruiting policy. The project's second segment, to be described in a forthcoming report, replicates the Hosek and Peterson studies (1985, 1990) and estimates additional models of enlistment decisions as a function of individual aptitudes and other characteristics.

While the NELS does not contain AFQT scores for respondents, it does contain extensive demographic information on those individuals and reports individuals' scores on math, science, and reading

tests. Several other studies have estimated AFQT scores from demographic characteristics using regression methodology (Grissmer et al., 1994, Orvis et al., 1996), but these were only able to explain less than half of the variance in individual test scores. Because the correlation between the AFQT and other composite tests is routinely high, we use respondents' scores on the tests in the NELS to estimate their AFQT scores rather than using regression to estimate AFQT scores.

Evidence from the National Assessment of Educational Progress (NAEP) indicates that youth aptitudes have improved between 1980 and 1992. This implies that we must account for this improvement when using the scores of NELS respondents in 1992 to predict their scores on the AFQT, which was normed on the youth population of 1980. That is, the raw score representing the 50th percentile in 1980 may represent a lower percentile in 1992. We use the NAEP math and reading results between 1980 and 1992 to adjust the 1992 NELS scores to make them comparable to the AFQT scores of 1980.

Our basic assumption is that a person who scored at some percentile on a NELS test would score at a similar percentile were the same population given a component of the NAEP or AFQT with comparable content. For example, if a person scored at the median on the NELS math test, he or she would score at approximately the median on the math portion of the NAEP. The NAEP provides comparisons of scores on math and reading tests over time. The steps involved in our procedure for both the math and reading tests are as follows:

1. Assign NELS respondents a 1992 NAEP percentile that is equivalent to their 1992 NELS percentile.
2. Assign NELS respondents a 1980 NAEP percentile that corresponds to the raw score for their 1992 NAEP percentile. This accounts for improvements in youth aptitudes.
3. Assign NELS respondents a 1980 AFQT percentile equivalent to their 1980 NAEP percentile, where the 1980 AFQT percentile is based on a subsample of the NLSY that matches the sampling scheme for the NELS.
4. Assign NELS respondents an adjusted AFQT percentile that is equal to the percentile their score attains in the 1980 NLSY AFQT norming sample.

This process not only yields an estimated AFQT score for each NELS respondent, but also provides some information with implications for recruiting policy. Specifically, we can compare the 1992 AFQT distribution we estimate to the 1980 AFQT test score distribution from the NLSY to get some initial insights into how the distribution has changed for various groups of interest to the recruiting community. When we compare the estimated score distribution of the NELS and the subsample of the NLSY matching the NELS sampling scheme, we find that in 1992 those designated AFQT CAT I–IIIA represent about 45 percent of high school seniors, while only 43 percent of high school seniors scored in those categories in 1980. In addition, while 7.7 percent of black seniors scored CAT I–IIIA in 1980, we estimate that approximately 19.7 percent of black seniors scored in this range in 1992. Hispanics also increased their representation in the upper portion of the distribution. These results suggest that a higher fraction of youths in 1992 scored in the ranges that qualified them for military enlistment than was true in 1980. This also implies that when the AFQT is renormed as planned using a random sample of the youth population collected in 1997—the NLSY-97—fewer individuals will qualify for enlistment than is currently true using the old AFQT norms. In other words, *ceteris paribus*, recruiters would be drawing from a smaller pool of eligible youths. Finally, the methodology used in our study suggests a way that AFQT scores renorming could be approximated on a regular basis between NLSY norming studies.

ACKNOWLEDGMENTS

This analysis has benefited from the input of many individuals. Marc Elliott provided help with statistical issues. Participants in the RAND Forces and Resources Policy (FRP) Center seminar provided valuable input, as did participants in a working seminar at DoD's Office of Accession Policy. Laurie McDonald provided strong programming support. Nora Wolverton assisted in the preparation of the manuscript. Jane Arabian, Beth Asch, Linda Curran, Steve Klein, Lee Lillard, LTC Michael Meese, Bruce Orvis, and Steve Sellman gave us important substantive comments. We also thank reviewers John Adams, Bruce Bloxom, Brian Stecher, and John Welsh for comments that substantially improved the report.

Chapter One

INTRODUCTION

Understanding the predictors of enlistment decisions helps recruiters and policymakers achieve their enlistment goals. In studies using the 1980 wave of the National Longitudinal Survey of Youth (NLSY), Hosek and Peterson (1985, 1990) showed that among those eligible to enlist, the probability that individuals enlist falls with increasing aptitude as measured by the Armed Forces Qualification Test (AFQT). In addition, they found that the decision to enlist was negatively related to an individual's wage rate and, for those who did not expect to obtain more education, positively related to mother's education. They also found that blacks were more likely than whites to enlist, that men were more likely than women to enlist, and that the effects of such background characteristics differed for high school seniors and graduates.

Since the collection of the data Hosek and Peterson used, changes have taken place that may have influenced individual enlistment probabilities. Among them are the following: the youth cohort is slightly smaller, the number of recruits needed has fallen sharply, the demographics of the youth population have changed such that a larger share of youths are minorities, youth aptitudes have risen, recruiter management has changed, a higher fraction of youths are attending college, the earnings of high school graduates have declined relative to college graduates, more recruits are female, and the military experienced the drawdown and engaged in the first war since Vietnam.

RAND has undertaken a project to estimate the predictors of individual enlistment decisions using a more recent data set, the National

Educational Longitudinal Study (NELS). This project (Kilburn and Klerman, forthcoming) both replicates the Hosek and Peterson studies (1985, 1990) and estimates additional models of enlistment decisions. AFQT is an important explanatory variable that is not included in the NELS data, so one of the first steps we must take is to estimate AFQT scores for the NELS participants. That is the purpose of the research reported in this paper. We also use the results of the estimation to draw some preliminary implications for recruiting policy.

Although NELS does not contain AFQT scores for respondents, it does contain extensive demographic information on those individuals and reports individuals' scores on math, science, and reading tests. Because the correlation between the AFQT and other cognitive ability tests such as these is routinely high, we use respondents' scores on the tests in NELS to estimate their AFQT scores.

Evidence from the National Assessment of Educational Progress (NAEP) (Grissmer et al., 1994, Koretz, 1992) and from High School and Beyond (Frankel et al., 1981), linked with NELS (Rasinski et al., 1993), indicates that youth aptitudes have improved between 1980 and 1992. Because AFQT scores are a percentile scale tied to the 1980 youth population norms, we must account for the change in ability in the youth population when using the 1992 scores of NELS respondents to estimate their scores on the AFQT. We use the NAEP math scores (from 1978, 1982, and 1992) and reading scores (from 1980 and 1992) to scale the 1992 NELS scores to be comparable to AFQT scores in the 1980 youth population—the score scale used by DoD for operational purposes.

In addition to estimating AFQT scores for the NELS respondents, we compare the distribution of estimated AFQT scores for high school seniors in the NELS to the estimated distribution of AFQT scores for high school seniors in the NLSY. If test scores have grown in the ways indicated by NAEP trends, we would expect the fraction of the NELS sample in upper AFQT categories to be higher than the fraction of the NLSY seniors.

This report has five chapters. Chapter Two describes the AFQT, its role in enlistment decision models, and the National Longitudinal Survey of Youth (NLSY), the sample used to norm the AFQT. Chapter

Three describes the NELS data and explains why we chose the method we use to estimate test scores. Chapter Four outlines in detail the methodology we use to estimate AFQT scores for NELS respondents. Chapter Five explores the implications of our estimation for broader issues in recruiting policy.

Chapter Two

THE AFQT AND ITS ROLE IN ENLISTMENT

AFQT scores have two important roles in the enlistment process. First, they indicate which individuals are eligible for enlistment. Eligibility for military enlistment is based on a combination of high school graduation status and test scores, age, citizenship, and dependency status, along with minimum health and moral requirements (see more detailed discussion in Kirby and Thie, 1996). The test score standards are mandated by law: Congress has stipulated that no recruits can come from the lowest 10 percent of the population distribution of AFQT scores and that only a quarter of recruits can come from the 10th to 30th percentiles (10 U.S.C. 520 and DoD Directive 1145.1). In fact, operational recruiting standards are typically much more stringent than these minimum congressional standards (see Eitelberg et al., 1984).

Second, as shown in Hosek and Peterson (1985, 1990), AFQT scores are an important determinant of individual enlistment decisions. Hosek and Peterson found a strong relationship between AFQT and enlistment probability: the probability that an individual enlisted dropped with rising AFQT scores.

Therefore, in order to estimate a model of the individual enlistment decision using more recent data, we also need to include AFQT or some close approximation to AFQT. AFQT scores are not available in any recent representative samples of potential recruits. We chose to use the NELS to estimate individual enlistment models for more recent cohorts for several reasons. First, the NELS contains demographic variables similar to those used as covariates by Hosek and Peterson. Second, the NELS reports which sample members en-

listed. Finally, the NELS contains cognitive test scores, which can be used to approximate AFQT scores. We now describe the AFQT in more detail.

THE ARMED FORCES QUALIFICATION TEST

The primary measure of aptitude for determining eligibility for admission into the Armed Services is an individual's score on the AFQT. The AFQT is designed to measure the trainability of potential recruits—more specifically, to identify individuals who are at high risk of not completing the initial training program (Eitelberg et al., 1984). The AFQT is a combination of scores from tests that are included in the Armed Service Vocational Aptitude Battery (ASVAB). The ASVAB is a ten-subtest battery administered to all military applicants. The test is designed to identify applicants who exceed minimum entry requirements and to match recruits to military occupations for which they are well suited. Since 1989, the AFQT has consisted of the sum of the standard scores[1] from the Arithmetic Reasoning and Math Knowledge subtests plus twice the sum of the standard scores on the Paragraph Comprehension and Word Knowledge subtests.[2]

The military divides AFQT percentiles into five categories that indicate subsets of the test score distribution (see Table 2.1). High school graduates in the top half of the AFQT distribution—CAT I–IIIA—are often referred to as the "high-quality" market, and individuals with these scores and educational status are considered the most desirable recruits, for their ability to succeed in and complete training.

As mentioned earlier, Congress mandates that no enlistees may come from the lowest 10 percentiles—CAT V—and that no more than 25 percent of enlistees can have scores between the 9th and 31st percentiles—CAT IV. Operational standards for recruiting often differ but do not fall below these legal standards for recruiting. Operational standards vary over time to reflect the needs of the service, the

---

[1]The ASVAB subtests are standardized to have a mean of 50 and standard deviation of 10 in the 1980 youth population.

[2]Before 1989, the AFQT score was equal to the sum of raw scores on the Word Knowledge, Paragraph Comprehension, and Arithmetic Reasoning subtests plus one-half the raw score on the Numerical Operations subtest.

Table 2.1

AFQT Percentiles and Categories

AFQT Percentile	AFQT Category
93–99	I
65–92	II
50–64	III–A
31–49	III–B
10–30	IV
1–9	V

ease or difficulty of recruiting due to labor market conditions, or other factors. For example, operational standards might require that all enlistees have a high school diploma or that recruits be restricted to CAT I–IIIB. In addition, recruiter incentives are designed in a way that will influence the mix of recruits. For instance, the incentives are often designed to encourage recruiters to enlist "high-quality" recruits.

THE NATIONAL LONGITUDINAL SURVEY OF YOUTH: THE CURRENT BASIS FOR AFQT

In 1980, the Department of Defense (DoD) administered the ASVAB to a nationally representative sample of youth as part of the NLSY. This effort—called the "Profile of American Youth" study (Bock and Moore, 1984, Office of the Assistant Secretary of Defense, 1982, and Maier and Sims, 1986)—is the only time that the ASVAB has been administered to a random sample of youths. Prior to the Profiles study, norms for the AFQT were based on the population of males on active duty on December 1, 1944, including both officers and enlisted personnel (Waters and Lindsley, 1996).

The original NLSY consists of a random sample of 6,111 youths who were age 13–20 in 1978 plus an oversample of 5,295 black, Hispanic, and economically disadvantaged youths who were not black or Hispanic as part of the National Longitudinal Survey of Youth (NLSY). The survey also included an oversample of 1,280 people who were enlisted in the military in 1979, but it dropped these individuals after

1985. In addition to taking the ASVAB, each respondent answered a broad range of questions in each year between 1979 and 1994.

Sponsored by both the Department of Labor (DoL) and DoD, the NLSY Profiles study had several purposes. First, it would allow DoD to identify percentile scores on the AFQT that were normed against a contemporary representative sample of the youth population. Second, it would permit DoD to measure the fraction of the youth population that would satisfy eligibility requirements and to examine differences in eligibility across demographic characteristics, geographic regions, or other factors. Third, it would facilitate the investigation of the relationship between vocational aptitudes and a large number of labor force and other outcomes (see, for example, O'Neill, 1990, Herrnstein and Murray, 1994 , Cameron and Heckman, 1993, Currie and Thomas, 1995, and others). A new, nationally representative sample of youth will be included in a second Profiles study as part of an NLSY study slated to begin in 1997. DoD and DoL will again jointly sponsor this survey and will administer the ASVAB as part of the study. The revised norms should be available sometime around the turn of the century.

Chapter Three

INFORMATION IN THE NELS AND APPROACHES TO IMPUTING TEST SCORES

As discussed in Chapter Two, knowing an individual's AFQT score is critical to making predictions about his or her enlistment decisions. Our study of individual enlistment decisions uses the NELS, a data set that does not report AFQT scores for respondents. We will now describe the NELS and then, given knowledge of that study's contents and structure, we will discuss alternative approaches to imputing AFQT scores for NELS respondents.

THE NATIONAL EDUCATION LONGITUDINAL STUDY

The National Education Longitudinal Study (NELS) follows a representative sample of individuals who were 8th graders in 1988, obtaining information on high school, postsecondary education, work, family formation, and background characteristics. The 1988 sample was selected using a two-stage probability strategy. In the first stage, approximately 1,000 public and private schools were selected from the universe of about 40,000 schools containing 8th graders. In the second stage, random samples of 24–26 students per school were selected. Also included in the sample are a parent, the school principal, and two teachers for each selected student. The study oversamples Hispanic and Asian students.

The study interviewed respondents in the base year (1988), a first follow-up (1990), a second follow-up (1992), and a third follow-up (1994). In each follow-up the school samples were freshened—a process that adds students to compensate for those dropping out, studying abroad, or emigrating—so that the sample remained representative of a random sample of students in a particular grade level.

So despite the fact that some students from each earlier wave of the study were no longer in school, the first follow-up is representative of students enrolled in 10th grade in the spring of 1990, and the second follow-up is representative of students enrolled in 12th grade in the spring of 1992. The third follow-up was not freshened.

Each interview includes a student questionnaire for individuals still in school, a dropout questionnaire for respondents no longer in school, a teacher questionnaire that asks teachers about specific respondents as well as class and school climate information, and a school questionnaire to obtain characteristics of the school. The student questionnaire collects information on family background, school activities, plans for the future, and other characteristics. The second follow-up also reports the respondent's score on cognitive tests in the areas of reading, math, science, and social science. These tests are unique to the NELS and were designed to measure the acquisition of aptitudes appropriate for the 12th grade. This follow-up also asks seniors if they have enlisted in the military.

The third follow-up surveys respondents two years after high school graduation. This questionnaire asks respondents to report on education, work, family formation, and other activities over this two-year period. We can identify which respondents enlisted during the period using both contemporaneous and retrospective questions. Hosek and Peterson (1985, 1990) distinguished between seniors and graduates. Using the questions in both the second and third follow-ups, we can also distinguish between individuals who enlisted while seniors and those who enlisted after graduating.[1]

APPROACHES TO IMPUTING TEST SCORES

Given the information available in the NELS, two major approaches could be used to estimate AFQT scores for NELS respondents. The first of these would use the multiple regression results from others'

[1]Although the NELS sample will allow us to study the enlistment behavior of one cohort of high school seniors, it does not provide a comprehensive view of enlistment behavior: individuals as old as 35 are eligible to enlist, as are those who never reached the 12th grade. However, most individuals who enlist do so at the ages at which we observe the NELS sample: nearly three-quarters of nonprior accessions are age 20 or younger (Office of the Assistant Secretary of Defense, 1994).

analyses that have regressed AFQT scores on demographic variables. We could use the demographic characteristics of members of the NELS to estimate AFQT scores based on the regression coefficients from other studies.

For example, Grissmer et al. (1994) use multivariate regression to estimate math and verbal scores of NLSY respondents using background characteristics such as age, race, gender, parents' education, family income, number of siblings, region of the country, and others. They are able to explain about 28 percent of the variance in math scores and 36 percent of the variance in verbal scores. They estimate similar regressions for the math and verbal scores in the NELS, with similar results: they explain about 28 percent of the total variance in the NELS math scores and 23 percent of the variance in NELS verbal scores. Note that when they estimate these models separately by race, they are able to explain substantially less of the variance for blacks and Hispanics—between 11 and 12 percent—than for whites—between 18 and 22 percent.

Neal and Johnson (1996) also use multivariate regression to estimate AFQT scores with the NLSY and obtain results close to those of Grissmer et al. (1994). Their model, which includes race, parents' education, parents' occupational status, number of siblings, and indicators of the learning environment at home and in school, explains up to 40 percent of the total variation in AFQT score.

Orvis et al. (1996) also use a similar approach. However, rather than estimating individuals' AFQT scores, they estimate the probability of an individual's scoring above the 50th percentile, in the CAT I–IIIA range. Orvis et al. obtained the actual AFQT scores of respondents in the 1984–1993 Youth Attitude Tracking Survey (YATS) who subsequently applied for enlistment before spring 1995. They estimate the probability that an individual scored CAT I–IIIA on the ASVAB as a function of characteristics reported by the respondent in the earlier YATS. The characteristics include the person's educational attainment, race, gender, parents' education, whether the person completed different types of courses in school, grade point average, region of the country, and other factors.

Multiple regression is an ideal methodology for accomplishing the objectives of Grissmer et al. (1994), Neal and Johnson (1996), and

Orvis et al. (1996). The former two papers were trying to understand how changes in test scores are related to changing demographics, and the latter was trying to weight the YATS to produce information on the high-quality recruiting market. But our objective in estimating AFQT scores is different. Our objective is to obtain the best estimate of individual AFQT scores given the information available to us in the NELS data set. We have instead chosen to implement a second strategy for imputing AFQT scores, one that uses the math and verbal test scores in the NELS data set rather than the demographic information.

We have three reasons to believe this strategy satisfies our objectives better than multiple regression. First, evidence suggests that using the math and verbal cognitive test scores available for NELS respondents to estimate of AFQT scores will explain more of the total variation in AFQT scores than regression does. While the demographic information in the NELS could account for as much as 40 percent of the variance in AFQT score regressions, other tests of math and verbal ability have been shown to account for as much as 70 percent or more of AFQT score variance. For example, AFQT scores correlate $r = 0.84$ with the composite of Verbal Reasoning and Numerical Ability from the DAT (Differential Aptitude Test), $r = 0.76$ with the composite of Mathematics Computation and Mathematics Concepts and Applications from the CAT (California Achievement Tests), and $r = 0.83$ with the composite of Reading Vocabulary and Reading Comprehension from the CAT (U.S. Military Entrance Processing Command, 1985). Bloxom et al. (1995) also reported that math items from the National Assessment of Educational Progress (NAEP) correlated $r = 0.85$ with an ASVAB math score. In addition, these correlations compare favorably with the alternate form reliabilities of the ASVAB subtests in AFQT that range from $r_{xx} = 0.80$ to $r_{xx} = 0.89$ (U.S. Military Entrance Processing Command, 1985). That is, AFQT correlates almost as highly with other similar composites (i.e., math and verbal) as its components correlate with alternate parallel forms. As a result, one might reasonably expect that estimates of AFQT derived from other reasonably parallel math and verbal composite scores will be nearly as accurate as the AFQT scores derived from alternate forms of ASVAB. Certainly they will be better approximations to AFQT scores than estimates derived from demographic information alone.

Second, while the multiple regression approach would allow us to explain changes across time in the mean of test scores, it would not allow us to reproduce shifts in all parts of the test score distribution. Results from the NAEP trend studies (Mullis et al., 1991, Hauser and Huang, 1996) show that test score gains were uneven across the test score distribution and that these patterns were different across demographic groups. Using the NELS test scores will better allow us to capture the actual changes across all parts of the distribution.

Third, test scores estimated from other test scores should provide better identification when we estimate enlistment models than would test scores estimated using demographic information alone. The reason we are estimating AFQT scores for NELS respondents is to use the scores as an explanatory variable in enlistment-decision models. These models will also include as explanatory variables a set of demographic characteristics such as mother's education, race, gender, region of the country, and others. Note that this list of additional explanatory variables is nearly identical to the demographic variables we would use to estimate AFQT using regression. This would result in an enlistment probability equation that was a function of demographic variables and an AFQT estimate that was itself a function of some of the same demographic variables. Unless we could find an instrument for estimating AFQT that was not in the set of demographic variables that predicted enlistment—and we could not discern any such variables—the effect of AFQT scores on enlistment would not be identified. An individual's score on a different test would contain information about that individual's likely score on the AFQT that was not captured by the demographic variables. Hence, an estimate of AFQT based on other test scores would serve as a valid instrument for AFQT score in our enlistment model.

We use the NELS math and verbal test scores to generate an approximation to individual AFQT scores using the methodology described in detail in the next chapter.

Chapter Four

METHODOLOGY

In Chapter Three we explained why we chose to estimate AFQT scores using NELS reading and math scores. In this chapter we describe this method in detail. We begin by discussing some of the issues we must consider in devising our methodology, and then we outline the steps of our test score imputing strategy.

ISSUES IN ESTIMATING AFQT SCORES FOR NELS RESPONDENTS

In considering the strategy that we outline below, it is important to keep in the forefront that our objective is to estimate for NELS respondents the AFQT scores they would have achieved if they had taken ASVAB. We have already noted our reasons for using the NELS math and verbal scores as the basis of our strategy. The question is, how can we use this information to estimate AFQT scores?

We begin with the assumption that the NELS math and verbal tests are sufficiently similar to the math and verbal components of AFQT that if the same sample of individuals took both sets of tests, their rank orderings on composite math and verbal scores would be identical across these tests, except for random error; i.e., we assume that an individual scoring at the 10th percentile of a given sample of individuals on a NELS math and verbal composite would score at the 10th percentile in the same sample of individuals on AFQT, and so on for all percentiles. Evidence supporting this assumption is discussed later in this chapter.

In fact, this basic assumption implies that if the NELS sample were equivalent demographically and equal in ability to the NLSY 1980 ASVAB norming sample, we would simply calculate a NELS 1992 composite math and verbal percentile score for the NELS respondents as an estimate of an AFQT percentile score.

However, three important facts keep us from following this strategy. First, the NELS and NLSY samples differ in terms of age. The NELS sample we are using is for 12th graders who were enrolled in school, whereas the NLSY ASVAB norming sample includes the 9,173 respondents aged 18 to 23, some of whom were in school and some not. In effect, the NLSY ASVAB norming sample includes individuals with higher levels of ability than will be found in the NELS sample, if only as the result of additional education (e.g., college degrees). Thus, all else equal, the percentiles associated with the same raw ability score would differ across these groups; i.e., percentile standing in NELS is not a good representation of percentile standing on AFQT.

Second, the demographics of our society have been changing in ways that influence the distributions of test scores (cf. Grissmer et al., 1994). To the extent that demographic characteristics are related to test scores and demographic characteristics have changed between 1980 and 1992, AFQT estimates based on percentiles on the 1992 NELS sample will be wrong. Consider an example to illustrate this point. Suppose the 1980 youth cohort contains one-third below-average youth, one-third average youth, and one-third above-average youth. The test score at the 50th percentile is the median for that population. Next suppose the 1992 youth cohort contains one-third average youth and two-thirds above-average youth—i.e., 1992 youth have higher relative raw scores on an ability test than the 1980 youth cohort. A youth scoring at the median of the 1992 cohort would be above the median when compared to the 1980 cohort (see Figure 1). Hence, estimating their ability using their standing compared to 1992 youth would not correctly identify their ability according to the AFQT norms, which are based on the ability of a 1980 population. Therefore, to the extent that changing demographics are related to overall changes in ability levels of the cohort, AFQT estimates based on percentiles on the 1992 NELS sample will be wrong.

Third, evidence from the National Assessment of Educational Progress (NAEP) indicates that youth aptitudes generally rose between 1980 and 1992. (Grissmer et al., 1994, Mullis et al., 1991). To the extent that this is true, a simple percentile on the 1992 NELS sample will underestimate AFQT scores. To the extent that improvements in test scores have been differential by demographic groups, we will underestimate scores for some more than for others. Consider the following example. Assume that a youth scoring 214 is at the 50th percentile in 1992. If raw scores had risen an average of 10 points since 1980, the score of 214 might place her as high as the 59th percentile on the 1980 scale, so that a percentile estimate based on 1992 raw scores would underestimate the 1980 score by 9 percentile points.

How can we take into account the changes in the youth population noted above in our estimates of AFQT? What we need is a test with a constant score scale that reaches across 1980 through 1992 and has been administered to nationally representative youth samples. If we had test score data on a representative sample of youth in 1980 and test score data *on the same score scale* for a representative sample of youth in 1992, changes in the score distributions from 1980 to 1992

RANDMR818-1

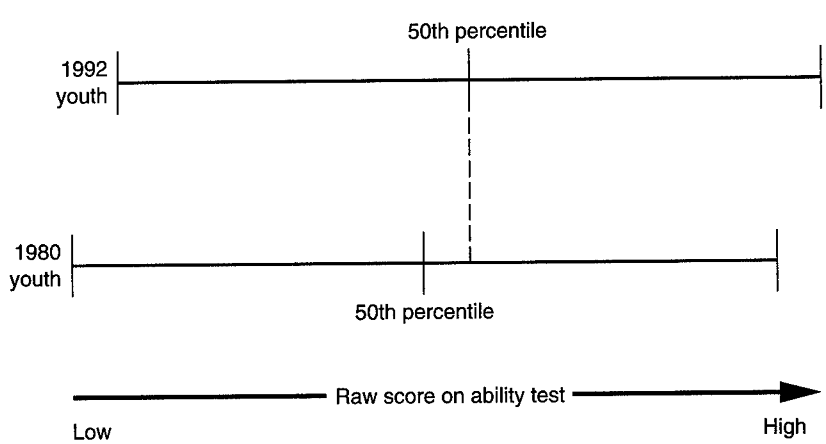

Figure 1—Changing Ability Affects Percentile Scores

would incorporate all of the effects of changes in demographics and schooling that occurred during that period and we would be able to calculate the percentile standing of a 1992 youth compared to the 1980 youth distribution—exactly the problem we are trying to address in estimating AFQT scores, on the 1980 score scale, for youth in 1992. The National Assessment of Educational Progress or NAEP trend study math and reading scale scores span the necessary time period (1980–1992).

The NAEP is a congressionally mandated program to monitor student performance. The core of NAEP is a set of assessments in reading, mathematics, science, and writing administered periodically to nationally representative samples of students. Prior to 1990, students aged 9, 13, and 17 were tested in reading, math, and science every five years. In 1990, Congress began requiring assessments every other year and added a writing assessment to the set of tests.

There were three other major changes in NAEP in 1983. First, the NAEP split into two assessments. The main assessment was maintained and an additional "trend" assessment was added (see Barron and Koretz, 1995/1996, for a discussion). We use the main assessment in this study.[1] Second, the sampling design was expanded to include grade-representative samples in addition to age-representative samples. As a result, NAEP has a 12th grade sample comparable to the 12th grade NELS sample. Third, the mode of administration was changed from a paced written and aural presentation, in which every respondent answered the same items, to a written-only presentation, in which different respondents answered different items. The small number of items that each respondent answers and the fact that different individuals answer different sets of items create a problem in assigning comparable scores to individuals. A plausible-value methodology has been used to provide estimates of individuals' scores on a comparable scale (Mislevy, Johnson, and Muraki, 1992).

Part of the charge to NAEP is to report on trends in academic progress across time. To do this, NAEP researchers have developed

[1]The NAEP trend assessment exhibits two shortcomings for the purposes of this study, primarily that it started in 1983 and that sample sizes of minority groups are extremely small.

scale scores for science, math, reading, and writing to be comparable across both time and age level of respondents. As a result, for example, a math scale score of, say, 250 in 1978 has the same underlying psychometric meaning as a score of 250 in 1992, regardless of the age or grade of the individual attaining it. This quality of NAEP trend scale scores plays an important role in our research in linking 1992 NELS scores to 1980 NLSY AFQT scores.

The NAEP study has published scale-score-to-percentile conversion tables for nationally representative samples of youth for those time periods. Using NAEP scores as a bridge, it is possible to estimate the links between the 1992 NELS math and verbal scores and the 1980 AFQT math and verbal score components. We describe in detail how we do this later in this chapter. Furthermore, because NAEP provides a constant scale score across time, it inherently captures the effects of changes in demographics and youth ability that we observe between 1980 and 1992.

However, this adds another largely untested but theoretically plausible assumption to our research: that the math and verbal (reading) components of AFQT, NAEP, and NELS are sufficiently alike that we can consider them to be randomly parallel tests. Two factors lead us to conclude that this assumption is reasonable. First, Bloxom et al. (1995) judged that the NAEP math test had sufficient overlap with ASVAB math tests to attempt a linking of the two. In their study, both sets of tests were administered to the same sample. However, they noted evidence of motivational differences in the test scores that caused them to suspect the results of the link they developed. Nonetheless, they conclude that it is sensible, on the basis of content, to link NAEP and AFQT. Second, the NELS math test has sufficient overlap with NAEP that researchers estimated NAEP math scale scores for NELS participants and included the NAEP math scale scores as part of the NELS data set, suggesting it is reasonable to link NELS with NAEP.

Unfortunately, we have little direct evidence from published reports linking the NELS reading test with the NAEP reading test or the NAEP reading test with the AFQT verbal test components. Based on one objective of NELS noted by Rock and Pollack (1995, p. 4), it seems reasonable to link NELS reading with NAEP reading:

> The tests should be sufficiently reliable to support change measurement, and be characterized by a sufficiently dominant underlying factor to support the Item Response Theory (IRT) model. This latter requirement is necessary to support the vertical equating between retestings as well as the cross-sectional linking with HS&B and NAEP, if desired.

As for linking NAEP reading with the verbal tests in AFQT, NAEP reading consists of asking students

> to read and answer questions based on a variety of materials, including informational passages, literary text, and documents . . . most questions were multiple choice and were designed to assess students' ability to locate specific information, make inferences based on information in two or more parts of a passage, or identify the main idea in a passage. For the most part, these questions measured students' ability to read either for specific information or for general understanding.[2]

In other words, the NAEP reading test is similar to the Paragraph Comprehension subtest that is included in the verbal component of AFQT. Although the verbal component of AFQT also includes a vocabulary test (Word Knowledge), it does not seem unreasonable to link NAEP reading with the verbal component of AFQT.

We do not undertake formal analysis of the content overlap of the measures from the three surveys. This is because our objective is to generate an instrument for NELS respondents' true AFQT score to be used as a regressor in the next phase of our analysis. The requirements for a valid instrument are simply that the instrument be correlated with the latent variable of interest and uncorrelated with the error term of the regression in which it will be used (Greene, 1993). Due to the high correlation found between the AFQT and a number of general tests, these conditions are likely to be satisfied. We do not claim to estimate with a high degree of certainty each NELS respondent's AFQT score. This should be kept in mind when interpreting the simulations in later chapters of this report.

---

[2]Mullis et al. (1991), p. 204.

A DETAILED STRATEGY FOR ESTIMATING AFQT SCORES FOR NELS RESPONDENTS

Figure 2 provides an overview of the steps we took in estimating AFQT scores for NELS participants, using NAEP scores as a bridge between the two:

1. **Develop a conversion table to match NELS standard scores to NAEP scale scores.** Since NELS participants have NAEP-equated math scale scores, this step was only necessary for the NELS-to-NAEP reading scores. We calculated percentile-to-standard-score tables using the 12th grade sample for each of the NELS and NAEP reading tests and created a table to convert NELS standard score to NAEP scale score by matching percentiles.

2. **Develop conversion tables to match NLSY standard scores to NAEP scale scores.** We calculated percentile-to-score conversion tables using the 17-year-old in-school sample for each of the NLSY and NAEP math and reading components. The NLSY math table converts the sum of standard scores on Math Knowledge (MK) and Arithmetic Reasoning (AR) to percentiles. The NLSY verbal table converts the standard scores on "2VE," twice the sum of standard scores on Paragraph Comprehension and Word Knowledge, to percentiles. We created NLSY to NAEP conversion tables by matching percentiles separately for the math and reading tests.

3. **Assign NLSY standard scores (AR+MK and 2VE) to NELS respondents.** Using the tables developed in steps 1 and 2, we assigned a standard score for the AR+MK and a standard score for the 2VE subtests to each NELS participant.

4. **Calculate an estimated AFQT percentile score for NELS respondents.** We calculated a sum of standard scores (AR+MK+2VE) for each NELS participant and used the official DoD conversion table to look up AFQT percentile scores.

The tables in Appendix A list the scores from each sample for each step outlined above.

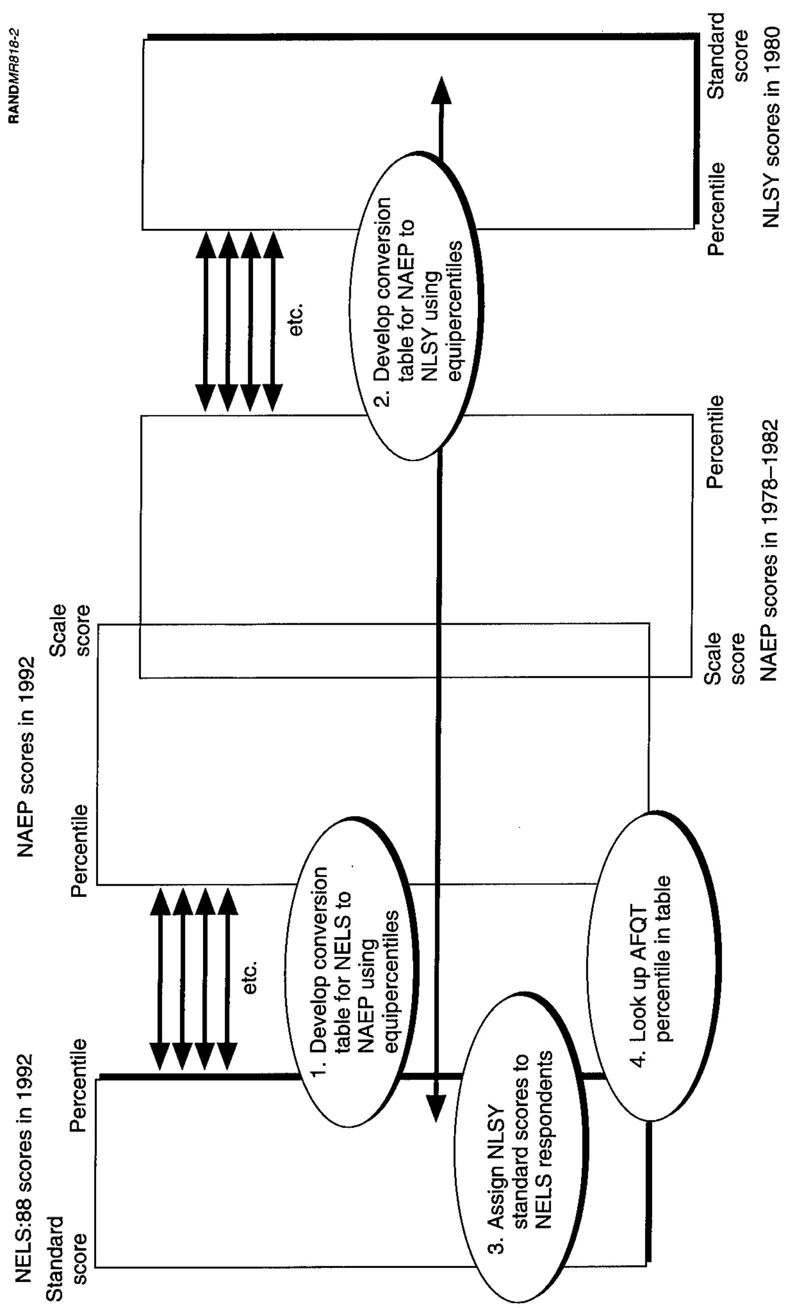

Figure 2—Estimating AFQT Scores for NELS Respondents

One important weakness of our method that is likely to influence the accuracy of our estimates is differences in the test content. For example, the NAEP verbal tests include paragraph comprehension and vocabulary components. In addition, the NAEP reading test includes constructed answers for some items—that is, not all items are multiple-choice as in the NLSY. Individually, this factor is not likely to create large errors in our estimates. Nonetheless, we do not suggest that the results on recruiting outcomes presented in the next chapter form the basis of policy without further analysis. The objective of this portion of the study was to generate an estimated AFQT to be included as an instrumental variable for unobserved true AFQT in enlistment regressions. We believe that the scale and rank orders of our estimated AFQT scores for NELS participants reflect AFQT scores well enough so that our estimates can fruitfully be used for this purpose.[3]

---

[3]Nevertheless, it is useful to consider error bounds on our AFQT estimates. We estimated standard errors for the NELS AFQT percentiles using an asymptotic sample approach (Serfling, 1980, section 2.6.2). These errors are for the best case under which there were no differences in test content and no errors in the equating procedure. These errors ranged from plus or minus 0.1 to plus or minus 1.3. An intermediate bound could be generated using a method for estimating standard errors for chained equating (Kolen and Brennan, 1995). But this method does not take into account the potential error due to differences in test content that would yield the upper bound error estimate.

Chapter Five

IMPLICATIONS FOR RECRUITING OUTCOMES

We undertook the analysis reported in this paper to support a study of individual enlistment decisions. In addition, however, the results of this analysis provide some insights into how the AFQT distribution may have changed since it was normed in 1980. This has important consequences for the military services. For example, continued success at recruiting high-quality youth may be an artifact of increasing ability levels that would raise the proportion of high-quality youth in the population when compared to 1980 norms.

First, let us compare the distribution of scores we estimated for the 1992 NELS and the subsample of the 1980 NLSY matching the NELS sampling scheme to examine how aptitudes on the AFQT changed over the period 1980 to 1992. Table 5.1 reports the estimated percentage of individuals in our NELS and NLSY subsample that scored in each AFQT category. Figure 3 shows the cumulative percentage in

Table 5.1

Estimated Percent of NLSY and NELS High School Seniors in Each AFQT Category

AFQT Category	NLSY High School Seniors (1980)	NELS High School Seniors (1992)
CAT I	6.0	5.3
CAT II	23.5	25.8
CAT IIIA	13.8	14.4
CAT IIIB	22.4	22.9
CAT IV	24.1	24.1
CAT V	10.1	7.6

each AFQT category. In 1992, approximately 45 percent of high school seniors scored AFQT CAT I–IIIA, while only 43 percent of high school seniors did so in 1980. Given the likely margin of error on these estimates, this is not likely to be a statistically significant difference. Also, whereas in 1980 about 10 percent of high school seniors scored CAT V—the range not eligible for enlistment by congressional mandate—we estimate that less than 8 percent scored in this range in 1992. This drop is relatively small and may not be statistically significant.

Second, we examine trends in test scores by race and gender. It is well known that black youth are more likely to enlist than white youth, despite their lower rates of eligibility.[1] Coupled with the fact that the NAEP results demonstrate that black test scores have risen dramatically over the last decade and a half, this may suggest that black enlistment could be on the rise, as their increasing test scores

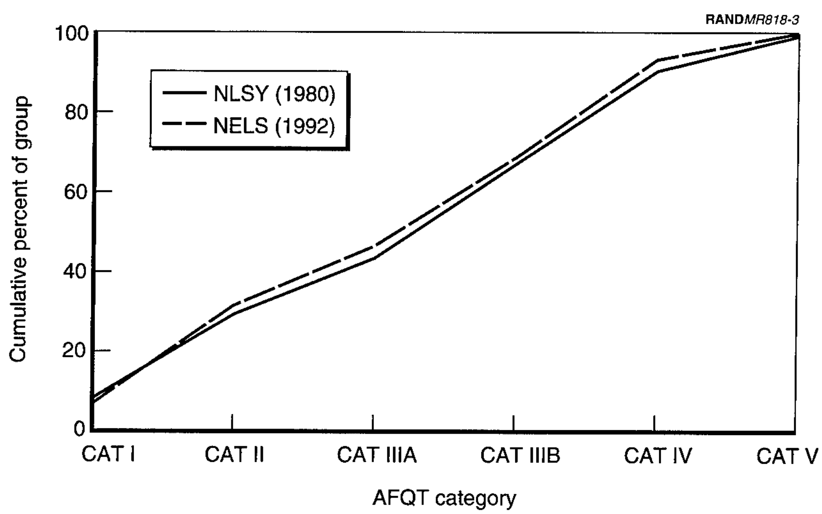

Figure 3—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category

---

[1]However, recent work (Kilburn and Klerman, forthcoming) shows that holding other background factors constant, there is no significant difference between the enlistment rates of black and white young men in the NELS.

imply more are eligible to enlist. As Table 5.2 shows, we estimate that more black high school seniors in the NELS were in the top three AFQT categories than was true in the NLSY. Figure 4 shows the substantial rise between 1980 and 1992 of the estimated fraction of black high school seniors who are eligible as well as the fraction who

Table 5.2

Estimated Percent of NLSY and NELS High School Seniors in Each AFQT Category, by Race/Ethnicity

	Black		Hispanic		White	
AFQT Category	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)
CAT I	0.3	0.6	1.3	1.6	7.5	6.7
CAT II	4.1	10.0	13.2	14.5	27.9	30.1
CAT IIIA	3.2	9.1	6.8	10.8	16.3	15.8
CAT IIIB	14.2	25.2	22.0	27.1	24.0	22.4
CAT IV	43.5	37.2	39.2	36.0	19.4	20.0
CAT V	34.7	17.9	17.5	10.1	5.0	5.2

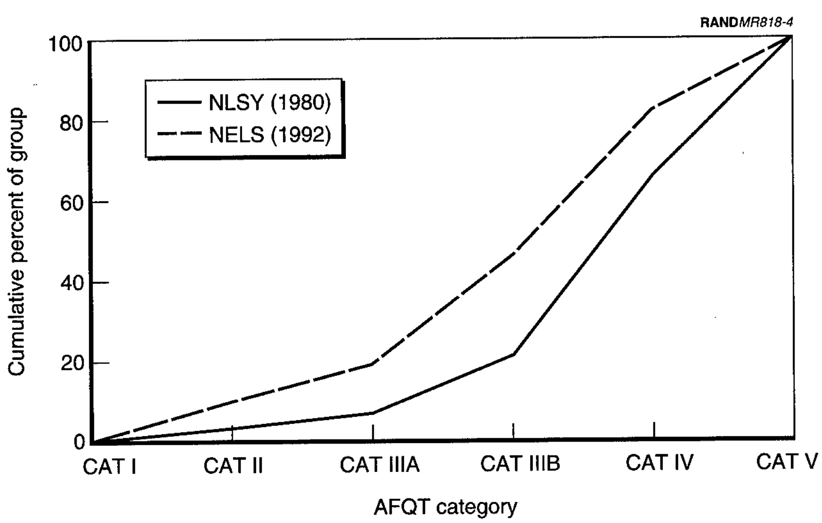

Figure 4—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category: Blacks

scored CAT I–IIIA. While less than 8 percent of black seniors scored CAT I–IIIA in 1980, about 20 percent of black seniors scored in this range in 1992.

Figure 5 shows that Hispanic high school students also increased their representation in the upper portion of the distribution: while 21.3 percent scored CAT I–IIIA in 1980, we estimate that by 1992, 26.9 percent scored in this range. There has been no meaningful change in the test scores of whites, as displayed in Figure 6.

Another group that posted large gains in test scores over the period 1980 to 1992 is high school females. While the scores for high school males grew slightly over the period—we estimate that the mean percentile score showed almost no change, rising from 47.2 to 48.0 between 1980 and 1992—the growth in the scores of high school females was substantially larger; we estimate that the mean percentile score rose from 45.0 to 49.4. We present estimates of the fraction of female and male respondents from the NLSY and NELS that scored in each AFQT category in Table 5.3. Figure 7 and Figure 8 report the estimated cumulative percent of high school seniors in each AFQT category for females and males, respectively, in our NLSY and NELS samples. In 1980, 40.6 percent of high school females scored in the CAT I–IIIA range. We estimate that by 1992, 46.3 percent of high school females did so. In contrast, 46.0 percent of high school males scored CAT I–IIIA in 1980, and we estimate that this was unchanged at 45.9 by 1992.

Table 5.3

Estimated Percent of NLSY and NELS High School Seniors in Each AFQT Category, by Gender

	Female		Male	
AFQT Category	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)
CAT I	3.9	5.1	8.1	6.0
CAT II	22.1	26.5	24.9	26.4
CAT IIIA	14.6	15.7	13.0	13.5
CAT IIIB	25.0	24.6	20.0	21.5
CAT IV	25.5	22.4	22.9	24.4
CAT V	9.0	5.8	11.2	8.2

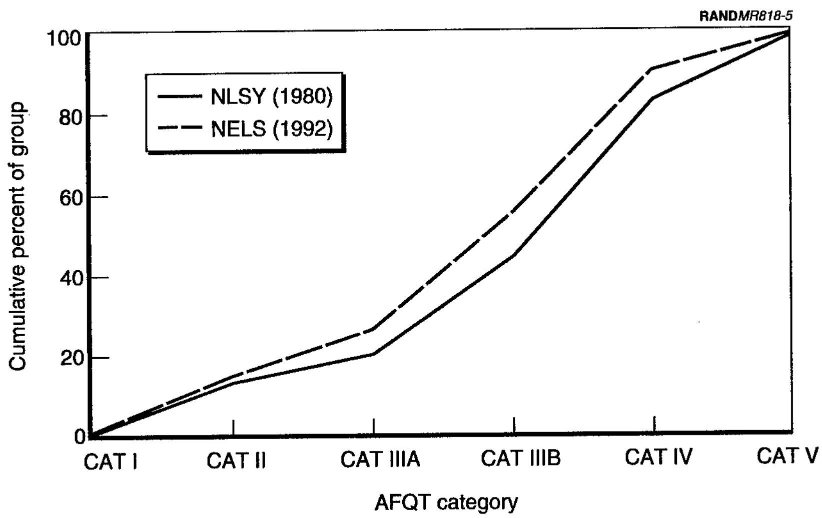

Figure 5—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category: Hispanics

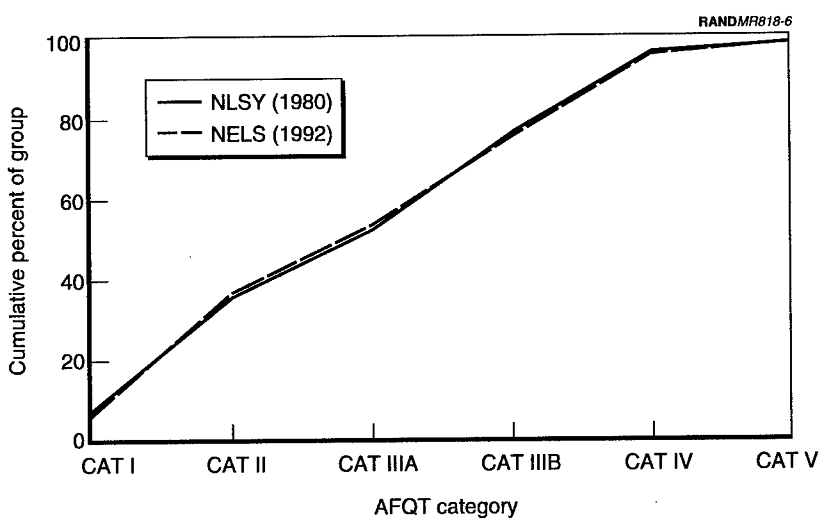

Figure 6—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category: Whites

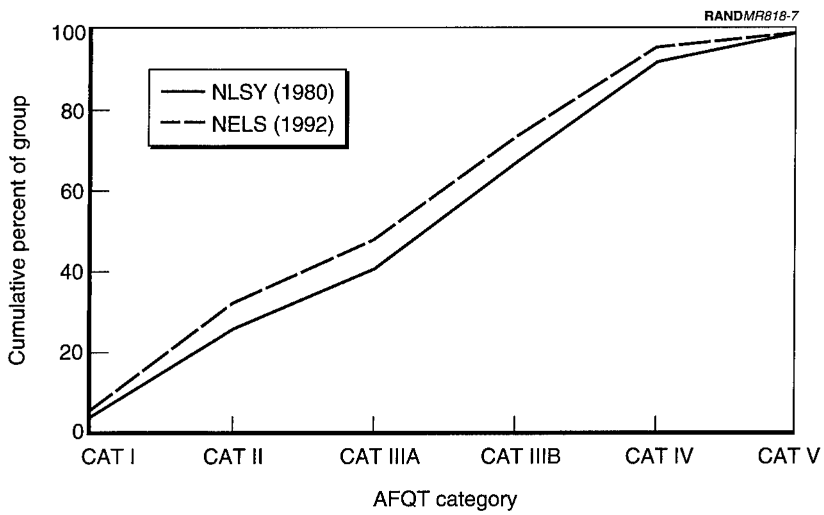

Figure 7—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category: Females

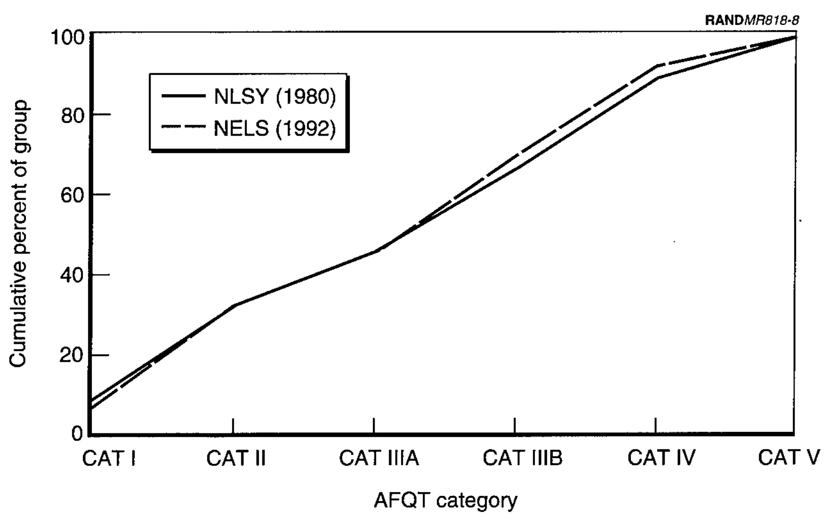

Figure 8—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category: Males

These results conform to the test score trends reported in other research (see, for example, Grissmer et al., 1994, and Mullis et al., 1991). However, while several references report test score trends by race and by gender, we do not know of published work that reports trends by race *and* gender. Given the relatively low propensity of women to enlist relative to men, in drawing implications of test score trends for recruiting outcomes it is essential to separate gains made by men and women. For instance, the gains in minority test performance reported in Table 4 would be less beneficial to recruiting if they derived primarily from females than if they derived mostly from males.

The estimates in Table 5.4 indicate that the test score gains within minority groups were not balanced between men and women. We estimate that females posted larger gains over the period than males. Figures 9–11 report the estimated cumulative percent of high school seniors in the NLSY and NELS in each AFQT category by race/ethnicity and gender. Black females posted a much larger gain in share of CAT I–IIIA scores between 1980 and 1992 than did black males (Figure 9). We estimate that black females increased their representation in the upper three categories from about 6 percent in 1980 to almost 23 percent in 1992. The representation of high school black males in the top three categories rose from 9.4 percent in 1980 to an estimated 16.0 percent in 1992. This indicates that black females contributed substantially more to gains in black test scores than did black males. Figure 10 shows that increases in scores for Hispanic females also account for more of the gains in test scores for Hispanics than do increases in scores due to Hispanic males. However, Hispanic males account for a much larger share of Hispanic test score growth than the score gains in black males explain out of the total growth in black test scores. We estimate that white women show modest AFQT score growth between 1980 and 1992 while the estimated fraction of white men in the top three categories appears to have declined slightly (see Figure 11), but these changes are not likely to be statistically significant.

Note that our findings are in contrast to other reports (Kageff and Laurence, 1994, for example) that the military should expect the pool of eligible youths to be shrinking rather than growing. These arguments often hinge upon the fact that minorities are increasing their fraction of the youth population. Such an argument does not take

Table 5.4

Percent of NLSY and NELS High School Seniors in Each AFQT Category, by Race/Ethnicity and Gender

	Female					
	Black		Hispanic		White	
AFQT Category	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)
CAT I	0.0	1.0	1.0	1.8	4.8	6.0
CAT II	4.3	12.6	7.5	10.7	26.6	30.3
CAT IIIA	1.7	9.3	6.5	11.2	17.7	17.1
CAT IIIB	13.2	26.0	25.2	27.3	27.2	24.1
CAT IV	48.1	34.5	46.5	38.9	19.5	18.6
CAT V	32.7	16.6	13.3	10.2	4.1	3.7
	Male					
	Black		Hispanic		White	
AFQT Category	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)	NLSY (1980)	NELS (1992)
CAT I	0.7	0.0	1.6	1.5	9.9	7.3
CAT II	4.0	7.2	18.7	18.2	29.2	29.9
CAT IIIA	4.7	8.8	7.1	10.4	15.0	14.5
CAT IIIB	15.1	24.3	18.8	26.8	21.0	20.6
CAT IV	39.0	40.1	32.3	33.1	19.3	21.2
CAT V	36.6	19.4	21.5	10.0	5.7	6.6

into account the growth in minorities' test scores, however, which is central to our analysis. Jaeger (1992) has noted similar trends in SAT scores, indicating that improvements in SAT scores have been greater within demographic groups than for the overall population. He also points out that the test score gains within groups have been masked by the rising fraction of SAT takers from the lower-scoring demographic groups.

In sum, although estimated aggregate trends in scores by race would suggest that more youths would be eligible for service in the race categories with the highest propensity to enlist, a less optimistic picture emerges when we break these trends down by gender. Most of the

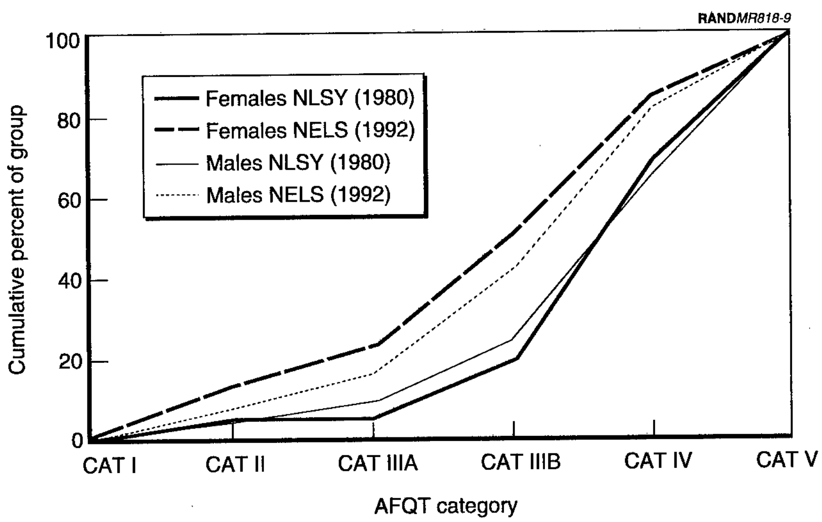

Figure 9—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category, by Gender: Blacks

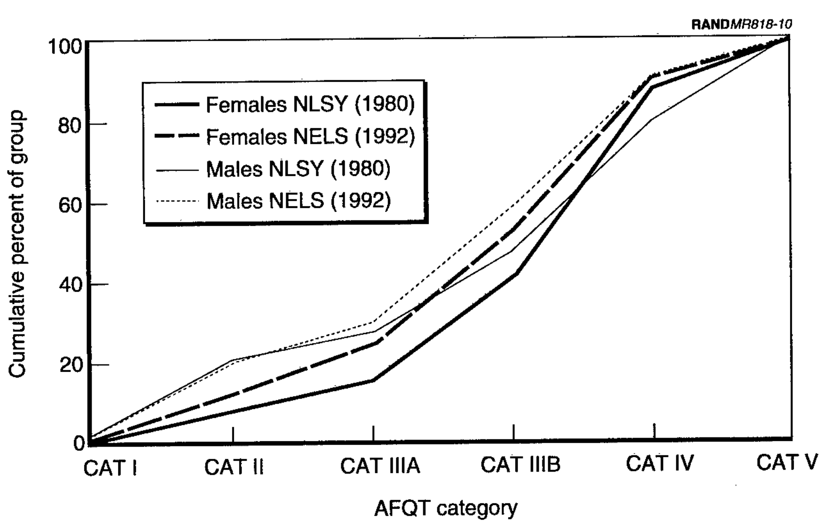

Figure 10—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category, by Gender: Hispanics

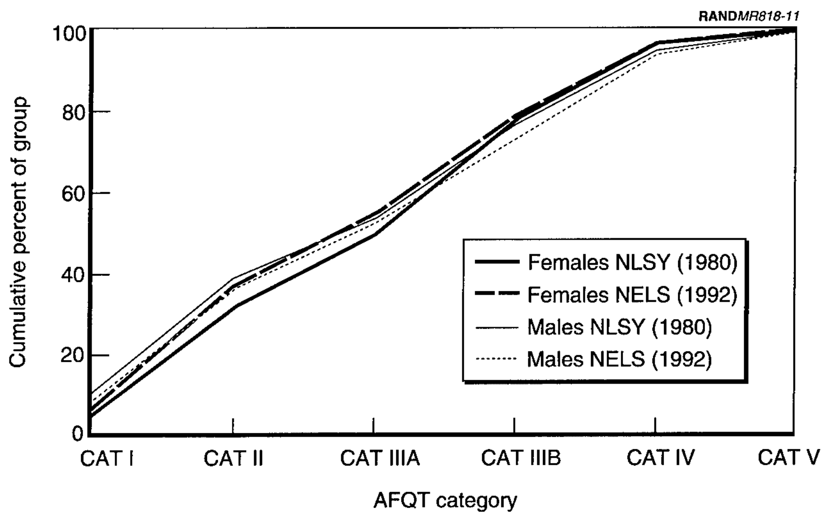

Figure 11—Cumulative Percent of NLSY and NELS Subsamples at Each AFQT Category, by Gender: Whites

improvements in blacks' estimated test scores and a large amount of the improvement in Hispanics' estimated test scores are due to increases in estimated female test scores. Since females have such a lower likelihood of joining the service than males do, the increases in eligibility due to rising test scores are unlikely to convert to appreciable increases in recruit supply. In Appendix B, we outline in more detail how one could relate test score trends to recruiting outcomes by estimating eligibility and using known propensities to enlist.

The implications of these estimated trends in test scores depend in part on the perspective one takes on the role of AFQT scores in recruit selection. There are at least two different perspectives. One perspective views the legislated minimum test scores as representing an absolute ability floor. Under this scenario, individuals with scores below this minimum would be unfit for military service. In that case, trends toward increasing test scores imply that the recruiting climate has, all else held equal, become easier: more 18-year-olds have scores above any given cutoff and thus will be eligible for enlistment.

A second perspective is to view the legislated minimum scores as relative scores. That is, military recruits could not be drawn from the lowest portion of the test score distribution regardless of absolute scores. The current congressional directives mandate that minimum test scores standards be measured *relative* to the percentiles in the population. In practice, DoD can only do this when the AFQT is renormed on a representative population—as was done in 1980 using the NLSY and will be done in the near future with the 1997 NLSY.

We could use estimates like those in this paper to project the effect of the renorming on recruiting. From this perspective, the recruiting situation is likely to get worse as test scores get better and the AFQT is renormed: some people who under the old norm would have met the old AFQT CAT cutoff will fail to do so after the renorming in 1997. In other words, after the renorming fewer individuals will qualify for enlistment, not because they have lower absolute abilities, but rather because the population distribution has increased and they have lower abilities relative to their contemporaries.

This study suggests there would be utility in exploring ways that AFQT scores could be tracked on a more regular basis than is possible with the NLSY norming studies. Given that the NAEP will continue to be administered for the foreseeable future, methods similar to ours could be developed to track the AFQT on a more regular basis between NLSY norming studies. To the extent that this would be valuable, the Department of Defense might want to explore sponsoring a formal norming of the NELS and NAEP whereby the same sample of students took both sets of tests.

Appendix A

STANDARD AND PERCENTILE TEST SCORE TABLES

Table A.1

Equated Standard and Percentile Scores on Math Tests

NELS NAEP-Equated Math Standard Score, 1992	NAEP Math Percentile Score, 1980	NLSY AR+MK Percentile Score, 1980	NLSY AR+MK Standard Score, 1980
220.15	1	1	68
229.35	2	2	69
235.20	3	3	70
239.55	4	4	72
243.10	5	5	72
245.80	6	6	73
248.35	7	7	74
250.75	8	8	75
253.00	9	9	75
255.05	10	10	76
257.00	11	11	76
258.90	12	12	77
260.60	13	13	77
262.25	14	14	78
263.80	15	15	79
265.30	16	16	79
266.70	17	17	80
268.05	18	18	80
269.30	19	19	80
270.55	20	20	81
271.75	21	21	81
272.85	22	22	82
274.00	23	23	82

Table A.1—Continued

NELS NAEP-Equated Math Standard Score, 1992	NAEP Math Percentile Score, 1980	NLSY AR+MK Percentile Score, 1980	NLSY AR+MK Standard Score, 1980
275.05	24	24	83
276.10	25	25	83
277.15	26	26	84
278.15	27	27	84
279.10	28	28	85
280.05	29	29	86
281.05	30	30	86
282.00	31	31	87
282.95	32	32	87
283.90	33	33	88
284.85	34	34	89
285.80	35	35	89
286.70	36	36	90
287.65	37	37	90
288.60	38	38	90
289.55	39	39	91
290.50	40	40	91
291.45	41	41	92
292.45	42	42	93
293.40	43	43	93
294.35	44	44	94
295.30	45	45	94
296.25	46	46	95
297.20	47	47	95
298.20	48	48	96
299.15	49	49	97
300.10	50	50	97
301.05	51	51	97
302.00	52	52	98
303.00	53	53	99
303.95	54	54	99
304.90	55	55	100
305.85	56	56	101
306.80	57	57	102
307.75	58	58	102
308.65	59	59	102
309.60	60	60	103
310.55	61	61	104

Table A.1—Continued

NELS NAEP-Equated Math Standard Score, 1992	NAEP Math Percentile Score, 1980	NLSY AR+MK Percentile Score, 1980	NLSY AR+MK Standard Score, 1980
311.45	62	62	105
312.40	63	63	105
313.30	64	64	106
314.20	65	65	107
315.15	66	66	108
316.05	67	67	108
316.95	68	68	109
317.90	69	69	109
318.80	70	70	110
319.70	71	71	111
320.60	72	72	112
321.55	73	73	113
322.45	74	74	113
323.45	75	75	114
324.45	76	76	115
325.45	77	77	116
326.45	78	78	116
327.50	79	79	117
328.55	80	80	117
329.70	81	81	119
330.85	82	82	120
332.10	83	83	120
333.40	84	84	121
334.70	85	85	122
336.10	86	86	123
337.60	87	87	123
339.20	88	88	125
340.85	89	89	125
342.65	90	90	127
344.55	91	91	127
346.55	92	92	128
348.75	93	93	129
351.00	94	94	130
353.45	95	95	131
357.00	96	96	131
361.35	97	97	132
367.20	98	98	133
376.40	99	99	134

Table A.2

Equated Standard and Percentile Scores on Reading Tests

NELS Reading Standard Score, 1992	NELS Reading Percentile Score, 1992	NAEP Reading Percentile Score, 1992	NAEP Reading Standard Score, Year Invariant	NAEP Reading Percentile Score, 1980	NLSY 2VE Percentile Score, 1980	NLSY 2VE Standard Score, 1980
30.71	1	1	195.9	.02	.02	48
31.67	2	2	203.1	.03	.03	54
32.56	3	3	208.6	.04	.04	56
33.25	4	4	213.0	.05	.05	58
33.87	5	5	217.1	.06	.06	60
34.39	6	6	220.8	.07	.07	62
34.89	7	7	224.3	.08	.08	64
35.51	8	8	227.6	.09	.09	66
36.04	9	9	230.6	.10	.10	66
36.56	10	10	233.4	.11	.11	68
37.10	11	11	238.4	.13	.13	72
251.7	12	12	236.0	.12	.12	70
253.2	13	13	240.7	.14	.14	72
38.56	14	14	242.8	.15	.15	74
38.96	15	15	244.8	.16	.16	74
39.45	16	16	246.7	.17	.17	74
39.92	17	17	248.5	.18	.18	76
40.35	18	18	250.2	.19	.19	78
40.80	19	19	253.2	.21	.21	80
41.24	20	20	251.7	.20	.20	78
41.69	21	21	254.7	.22	.22	80
42.09	22	22	256.1	.23	.23	82
42.44	23	23	257.4	.24	.24	84
42.87	24	24	258.7	.25	.25	84
43.30	25	25	260.0	.26	.26	84
43.68	26	26	262.4	.28	.28	88
44.10	27	27	261.2	.27	.27	86
44.43	28	28	263.6	.29	.29	88
44.84	29	29	264.7	.30	.30	90
45.24	30	30	265.9	.31	.31	90
45.65	31	31	267.0	.32	.32	90
45.96	32	32	268.1	.33	.33	90
46.30	33	33	270.4	.35	.35	92
46.75	34	34	269.3	.34	.34	92
47.14	35	35	271.5	.36	.36	92
47.49	36	36	272.6	.37	.37	94
47.89	37	37	273.8	.38	.38	94
48.20	38	38	274.9	.39	.39	94

Table A.2—Continued

NELS Reading Standard Score, 1992	NELS Reading Percentile Score, 1992	NAEP Reading Percentile Score, 1992	NAEP Reading Standard Score, Year Invariant	NAEP Reading Percentile Score, 1980	NLSY 2VE Percentile Score, 1980	NLSY 2VE Standard Score, 1980
48.63	39	39	276.0	.40	.40	96
48.98	40	40	277.2	.41	.41	96
49.32	41	41	278.3	.42	.42	96
49.71	42	42	279.5	.43	.43	98
50.00	43	43	280.6	.44	.44	98
50.33	44	44	281.7	.45	.45	98
50.65	45	45	282.9	.46	.46	100
50.94	46	46	284.1	.47	.47	100
51.29	47	47	285.2	.48	.48	100
51.66	48	48	286.4	.49	.49	100
52.00	49	49	287.5	.50	.50	102
52.29	50	50	288.6	.51	.51	102
52.56	51	51	289.8	.52	.52	102
52.92	52	52	290.9	.53	.53	102
53.20	53	53	292.1	.54	.54	102
53.51	54	54	293.2	.55	.55	104
53.79	55	55	294.3	.56	.56	104
54.01	56	56	295.4	.57	.57	104
54.34	57	57	296.5	.58	.58	104
54.67	58	58	297.6	.59	.59	106
54.87	59	59	298.7	.60	.600	106
55.15	60	60	299.8	.61	.61	106
55.44	61	61	300.9	.62	.62	106
55.68	62	62	301.9	.63	.63	106
56.04	63	63	303.0	.64	.64	108
56.28	64	64	304.0	.65	.65	108
56.52	65	65	305.1	.66	.66	108
56.80	66	66	306.1	.67	.67	108
57.05	67	67	307.2	.68	.68	108
57.30	68	68	308.2	.69	.69	108
57.58	69	69	309.3	.70	.70	108
57.79	70	70	310.3	.71	.71	108
58.03	71	71	311.4	.72	.72	110
58.31	72	72	312.4	.73	.73	110
58.58	73	73	313.5	.74	.74	110
58.84	74	74	314.6	.75	.75	110
59.08	75	75	315.7	.76	.76	110
59.33	76	76	316.9	.77	.77	112

Table A.2—Continued

NELS Reading Standard Score, 1992	NELS Reading Percentile Score, 1992	NAEP Reading Percentile Score, 1992	NAEP Reading Standard Score, Year Invariant	NAEP Reading Percentile Score, 1980	NLSY 2VE Percentile Score, 1980	NLSY 2VE Standard Score, 1980
59.58	77	77	318.1	.78	.78	112
59.88	78	78	319.3	.79	.79	112
60.17	79	79	319.95	.795	.795	112
60.43	80	80	320.6	.80	.80	113
60.61	81	81	321.9	.81	.81	114
60.88	82	82	323.3	.82	.82	114
61.18	83	83	324.7	.83	.83	114
61.51	84	84	326.2	.84	.84	114
61.77	85	85	327.8	.85	.85	114
62.03	86	86	329.5	.86	.86	116
62.31	87	87	340.45	.865	.865	116
62.50	88	88	331.4	.87	.87	116
62.85	89	89	333.3	.88	.88	116
63.29	90	90	335.3	.89	.89	116
63.61	91	91	337.5	.90	.90	118
63.95	92	92	339.8	.91	.91	118
64.31	93	93	342.3	.92	.92	118
64.65	94	94	343.65	.925	.925	118
64.91	95	95	345.0	.93	.93	119
65.35	96	96	347.8	.94	.94	120
65.85	97	97	350.9	.95	.95	120
66.40	98	98	355.3	.96	.96	120
67.02	99	99	358.05	.965	.965	122
—	—	—	360.8	.97	97	122
—	—	—	368.0	.98	.98	122
—	—	—	379.4	.99	.99	124
—	—	—	380.0	.99	.99	124
—	—	—	—	—	—	124
—	—	—	—	—	—	124

Appendix B

RELATING TEST SCORE ESTIMATES TO RECRUITING OUTCOMES

This appendix illustrates how one might relate estimates of changes in test scores to recruiting outcomes. The figures used for this example are not precise, but rather are intended to represent general principles.

As mentioned above, test scores are one of the primary means used to determine eligibility for enlistment. In 1992, over 99 percent of individuals who enlisted scored in AFQT categories I–IIIB, compared to less than an estimated 70 percent of the civilian youth population (Office of the Assistant Secretary of Defense (Personnel and Readiness), 1993). For this illustration, we consider the minimum criterion for eligibility for enlistment to be scoring in AFQT CAT I–IIIB. We show how changes in the fraction of different demographic groups scoring in CAT I–IIIB relate to the number of recruits, all else held constant.

The number of recruits is related to the fraction of individuals in CAT I–IIIB as follows. Let R_{jt} represent the number of recruits from group j in year t, N_{jt} represent the number in group j in the population in year t, and $P_{jt}(E)$ the probability that an individual from group j enlists in time t. The number of recruits from group j in year t equals

$$R_{jt} = P_{jt}(E)N_{jt} \; . \tag{1}$$

Since only eligible individuals can enlist, we can relate the probability of enlistment to the probability of eligibility as follows:

$$P_{jt}(E) = P_{jt}(E \mid G)P_{jt}(G),$$

where $P_{jt}(G)$ is the probability that an individual from group j is eligible to enlist in time t, and $P_{jt}(E \mid G)$ is the probability that an individual from group j enlists in time t given that the individual is eligible. Since, in this example, individuals are eligible if they score CAT I–IIIB on the AFQT, we can rewrite equation (1) to relate the number of recruits to the probability of scoring CAT I–IIIB:

$$R_{jt} = P_{jt}(E \mid G)P_{jt}(G)N_{jt}.$$

The change in the number of recruits from group j, ΔR_{jt}, given a change in the fraction of group j scoring in CAT I–IIIB, $\Delta P_{jt}(G)$, can be written

$$\Delta R_{jt} = P_{jt}(E \mid G)\left[\Delta P_{jt}(G)\right]N_{jt}. \quad (2)$$

Note that this relation assumes that all other factors related to recruiting such as propensity, civilian labor market alternatives, and others, are being held constant.

We use equation (2) to relate our estimated changes in the proportion of individuals in different demographic groups scoring CAT I–IIIB to recruiting yields. We ask the following question: What would the recruiting yield have been in 1992 had group j scored at 1980 levels rather than at 1992 levels? As discussed above, every demographic group except for white males improved its test scores over the period 1980 to 1992. Hence, we are examining what would have been recruiting yields, *ceteris paribus*, had test scores not improved. The same methodology could be used to ask what the recruiting yield would be if eligibility was based on 1992 AFQT norms—leaving fewer people in CAT I–IIIB—rather than 1980 AFQT norms given appropriate data. A similar approach would also indicate the change in the number of recruits due to population growth—changes in N_{jt}—or changes in enlistment probabilities given eligibility—changes in $P_{jt}(E \mid G)$.

We obtain rough estimates of R_{j92} from Office of the Assistant Secretary of Defense (Personnel and Readiness) (1993). Table B.1 and Table B.2 show the value of R_{j92} for various demographic groups. We used our estimate of the fraction of individuals in CAT I–IIIB from

the NELS as an estimate of the probability that individuals in group j would be eligible in 1992 or $P_{j92}(G)$. Table B.1 and Table B.2 also list these values by demographic group. While this is an estimate of the number of individuals in 12th grade in 1992 who are eligible rather than the number of youths who are eligible, we can interpret our equation (2) as the relation between the number of individuals from that subsample of the population who are eligible and the number of them who ever enlist, taking R_{j92} as an approximation of the number of 12th graders who ever enlist.[1]

Table B.1

Parameter Values for 18-Year-Old Males and Females

	18-Year-Old Males	18-Year-Old Females
R_{j92}	54,998	8,684
N_{jt}	1,642,628	1,594,452
$P_{jt}(E)$	.03348	.005446
$P_{j92}(G)$	.6737	.7184
$P_{jt}(E \mid G)$	.0497	.00758

Table B.2

Parameter Values by Race/Gender

	White Males	White Females	Black Males	Black Females	Hispanic Males
R_{j92}	126,578	19,329	26,250	7,221	13,141
N_{jt}	8,440,591	8,667,780	1,651,810	1,833,999	1,417,519
$P_{jt}(E)$	.0150	.0022	.0159	.0039	.00927
$P_{j92}(G)$	.7221	.7765	.4044	.4891	.5690
$P_{jt}(E \mid G)$	.0208	.0028	.0393	.0080	.0163

[1]If patterns of enlistment by age do not vary substantially over time, R_{j92} will approximate the number of individuals from the 92 "cohort" who ever enlist, since it contains the number of individuals from several other cohorts who enlisted at different ages.

Office of the Assistant Secretary of Defense (Personnel and Readiness) (1993) also lists the number of individuals from each group j in the population. We use this as our estimate of N_{j92}. Finally, we approximate $P_{jt}(E \mid G)$ by dividing

$$\frac{R_{j92}}{N_{j92}} = P_{j92}(E)$$

by our estimate of $P_{j92}(G)$. Again, our estimates for these values are presented in Tables B.1 and B.2.

We estimate $P_{j80}(G)$ by the fraction of group j in CAT I–IIIB in the 1980 NLSY 12th grade sample described above. Hence, the change in eligibility between 1980 and 1992 is

$$\Delta P_j(G) = P_{j92}(G) - P_{j80}(G).$$

ΔR_j is the difference between the number of recruits obtained given 1992 test scores and eligibility rates, R_{j92}, and the number of recruits that would have been obtained given 1980 test scores and eligibility rates, R_{j80}. Restated in equation form, this is

$$DR_j = R_{j92} - R_{j80} = P_{j92}(E \mid G)\left[P_{j92}(G) - P_{j80}(G)\right]N_{j82}.$$

Tables B.3 and B.4 display the results of these calculations.

Table B.3 shows that for 18-year-olds, the fraction of individuals eligible for enlistment as a result of scoring CAT I–IIIB on the AFQT rose between 1980 and 1992 for both males and females. The increase was more dramatic for females, as shown by $\Delta P_j(G)$. Had the ability levels of 18-year-olds stayed at 1980 levels, and all other factors remained constant, the military would have recruited 1,175 fewer 18-year-old males and 761 fewer 18-year-old females. These numbers represent 2.1 percent and 8.8 percent of 18-year-old male and 18-year-old female recruits, respectively.

Table B.4 reports the same calculations by race and gender groups. Black females had the largest gain in fraction of individuals eligible: the fraction scoring in CAT I–IIIB rose from .19 in 1980 to .49 in 1992.

Table B.3

Estimates of Change in Number of Recruits Resulting from Changing Test Scores, by Gender for 18-Year-Olds

	18-Year-Old Males	18-Year-Old Females
$P_{j92}(E \mid G)$	.0497	.00758
$P_{j92}(G)$	.6737	.7184
$P_{j80}(G)$	.6593	.6554
$\Delta P_j(G)$	.0144	.0630
R_{j92}	54,998	8,684
ΔR_j	1,175	761
% ΔR_j (from 1992)	2.1%	8.8%

Table B.4

Estimates of Change in Number of Recruits Resulting from Changing Test Scores, by Race/Gender Group

	White Males	White Females	Black Males	Black Females	Hispanic Males
$P_{j92}(E \mid G)$	.0204	.0028	.0375	.0079	.0152
$P_{j92}(G)$	.7221	.7765	.4044	.4891	.5690
$P_{j80}(G)$	.7496	.7635	.2445	.1913	.4624
$\Delta P_j(G)$	–.0275	.0130	.1599	.2978	.1066
R_{j92}	126,578	19,329	26,250	7,221	13,141
ΔR_j	–4,822	319	10,385	4,355	2,462
% ΔR_j	–3.81%	1.65%	39.6%	60.3%	18.7%

This implies that about 60 percent fewer or around 4,355 fewer black females would have enlisted, *ceteris paribus,* had test scores not risen the way they did between 1980 and 1992. Black males also posted large gains in the number eligible, with an estimated 40 percent fewer enlisting had test score trends not occurred. Note that even

though the fraction of black males eligible for enlistment did not grow as much as the fraction of black females, the difference in the number of enlistments for black males—10,385—is the largest of any group. This is because black males have such a high probability of enlisting given eligibility ($P_{jt}(E \mid G)$) relative to other groups. The fraction of Hispanic males who were eligible also grew considerably over the period, contributing to an estimate of 2,462 or 19 percent fewer Hispanic male enlistments without test score gains. Changes in eligibility for white females was positive but relatively small—a gain of 2 percent more eligible in 1992 than in 1980. The fraction of white males estimated to be in CAT I–IIIB actually declined by almost 4 percent.

If we sum over the changes from the five demographic groups that cross race and gender, we obtain a rough estimate of the total change in recruits due to the difference in test scores between 1992 and 1980. That is,

$$\sum_j \Delta R_j = \text{Total Change in Number of Recruits.}$$

We estimate that the total change in recruits due to the changes in scores within the five race/gender groups in Table B.4 is approximately 8.5 percent. Since these groups made up about 95 percent of recruits in 1992, this is a close approximation to the total estimated change. Table B.5 shows the percentage that each group contributed to the total change in recruiting, $\%\Delta S_j$. This is computed as the percentage change in the number of recruits from group j, $\%\Delta R_j$, weighted by the fraction of total recruits that come from group j, F_j. This table shows that the rise in test scores of black males contributed over half of the total difference in the number of recruits, accounting for an estimated 5.5 percent difference. The rise in test scores of black females and Hispanic males accounts for the bulk of the remainder of the difference. The estimated decline in the test scores of white males accounts for a decline in the number of recruits of about 2.5 percent.

While these calculations are imprecise "back-of-the-envelope" estimates, they illustrate how estimates from test score trend studies could be related to recruiting outcomes.

Table B.5

Estimates of Total Change in Number of Recruits Resulting from Changing Test Scores, by Demographic Group

	White Males	White Females	Black Males	Black Females	Hispanic Males
$\% \Delta R_j$	−3.81%	1.65%	39.6%	60.3%	18.7%
F_j	.639	.096	.130	.036	.065
$\% \Delta S_j$	−2.43%	.16%	5.14%	2.16%	1.22%

BIBLIOGRAPHY

Barron, Sheila I., and Daniel M. Koretz, "An Evaluation of the Robustness of the National Assessment of Education Progress Trend Estimates for Racial-Ethnic Subgroups," *Educational Assessment,* Vol. 3, No. 3, 1995/1996, pp. 209–248.

Bock, R. Darrell, and Elsie G. J. Moore, *Profile of American Youth, Demographic Influences on ASVAB Test Performance,* Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics), 1984.

Bloxom, Bruce, Peter J. Pashley, W. Alan Nicewander, and Duanli Yan, "Linking to a Large-Scale Assessment: An Empirical Evaluation," *Journal of Educational and Behavioral Statistics,* Vol. 20, No. 1, Spring 1995, pp. 1–26.

Cameron, Stephen V., and James J. Heckman, "The Nonequivalence of High School Equivalents," *Journal of Labor Economics,* Vol. 11, No. 1, 1993, Part 1, pp. 1–47.

Currie, Janet, and Duncan Thomas, "Race, Children's Cognitive Achievement and the Bell Curve," Santa Monica, CA: RAND Labor and Population Program Working Paper 95-19-1, DRU-1178-1-NICHD, 1995.

Eitelberg, Mark J., *Manpower for Military Occupations,* Office of the Assistant Secretary of Defense (Force Management and Personnel), Washington, D.C., April 1988.

Eitelberg, Mark J., Janice H. Laurence, Brian K. Waters, with Linda S. Perelman, *Screening for Service: Aptitude and Education Criteria*

For Military Entry, Office of Assistant Secretary of Defense (Manpower, Installations and Logistics), Washington, D.C., September 1984.

Frankel, M.R., L. Kohnke, D. Buonanno, and R. Tourangeau, *HS&B Base Year Sample Design Report,* Chicago: NORC, 1981.

Greene, William H., *Econometric Analysis,* 2nd ed., New York: Macmillan Publishing Company, 1993.

Grissmer, David W., Sheila Nataraj Kirby, Mark Berends, and Stephanie Williamson, *Student Achievement and the Changing American Family,* Santa Monica, CA: RAND, MR-488-LE, RAND, 1994.

Hauser, Robert M., and Min-Hsiung Huang, "Trends in Black-White Test Score Differentials," Presentation at Institute for Research on Poverty, University of Wisconsin, Summer Workshop, June 1996.

Herrnstein, Richard J., and Charles Murray, *The Bell Curve,* New York: Free Press, 1994.

Hosek, James R., and Christine E. Peterson, *Enlistment Decisions of Young Men,* Santa Monica, CA: RAND, R-3238-MIL, July 1985.

Hosek, James R., and Christine E. Peterson, *Serving Her Country: An Analysis of Women's Enlistment,* Santa Monica, CA: RAND, R-3853-FMP, January 1990.

Jaeger, R. M., "World Class Standards, Choice and Privatization: Weak Measurement Serving Presumptive Policy, " *Phi Delta Kappan,* Vol. 74, 1992, pp. 118–128.

Kageff, Linda L., and Janice H. Laurence, "Test Score Trends and the Recruit Quality Queue," in Mark J. Eitelberg and Stephen L. Mehay (eds.), *Marching Toward the 21st Century,* Westport, CT: Greenwood Press, 1994.

Kilburn, M. Rebecca, and Jacob A. Klerman, *Enlistment Decisions in the 1990s: Evidence from Individual-Level Data,* Santa Monica, CA: RAND, MR-944-OSD/A, forthcoming.

Kirby, Sheila Nataraj, and Harry J. Thie, *Enlisted Personnel Management: An Historical Perspective,* Santa Monica, CA: RAND, MR-755.0-OSD, May 1996.

Kolen, Michael J., and Robert L. Brennan, *Test Equating: Methods and Practices,* New York: Springer-Verlag, 1995.

Koretz, Daniel, "What Happened to Test Scores, and Why?" *Educational Measurement: Issues and Practice,* Winter 1992, pp. 7–11.

Maier, Milton H., *Military Aptitude Testing: The Past Fifty Years,* DMDC Technical Report 93-007, Monterey, CA: Personnel Testing Division, Defense Manpower Data Center, June 1993.

Maier, M. H., and W. H. Sims, *The ASVAB Score Scales: 1980 and World War II,* Alexandria, VA: Center for Naval Analyses, CNA CNR 116, July 1986.

Mislevy, R.J., E.G. Johnson, and E. Muraki, "Scaling Procedures in NAEP," *Journal of Educational Statistics,* Vol. 17, 1992, pp. 131–154.

Mullis, Ina V.S., John A. Dossey, Mary A. Foertsch, Lee R. Jones, and Claudia A. Gentile, *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969–70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; Writing, 1984 to 1990,* U.S. Department of Education, NCES 21-T-01, November 1991.

Neal, Derek A., and William R. Johnson, "The Role of Pre-Market Factors in Black-White Wage Differences," *Journal of Political Economy,* Vol. 104, No. 5, October 1996, pp. 869–895.

O'Neill, June, "The Role of Human Capital in Earnings Differences Between Black and White Men," *Journal of Economic Perspectives,* Vol. 4, No. 4, 1990, pp. 25–45.

Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), *Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery,* Washington, D.C., March 1982.

Office of the Assistant Secretary of Defense (Personnel and Readiness), *Population Representation in the Military Services, Fiscal Year 1992,* Washington, D.C., October 1993.

Office of the Assistant Secretary of Defense (Force Management Policy), *Population Representation in the Military Services, Fiscal Year 1993,* Washington, D.C., November 1994.

Office of the Assistant Secretary of Defense (Force Management Policy), *Population Representation in the Military Services, Fiscal Year 1994,* Washington, D.C., December 1995.

Orvis, Bruce R., Narayan Sastry, and Laurie L. McDonald, *Military Recruiting Outlook: Recent Trends in Enlistment Propensity and Conversion of Potential Enlisted Supply,* Santa Monica, CA: RAND, MR-677-A/OSD, 1996.

Rasinski, Kenneth A., Steven J. Ingels, Donald A. Rock, Judith M. Pollack, *America's High School Sophomores: A Ten Year Comparison,* Washington, D.C.: U.S. Government Printing Office, NCES 93-087, June 1993.

Rock, Donald A., Judith M. Pollack, and Peggy Quinn, *Psychometric Report for the NELS: 88 Base Year Through Second Follow-Up,* Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education, NCES 95-382, August 1995.

Serfling, Robert J., *Approximation Theorems of Mathematical Statistics,* New York: John Wiley & Sons, 1980.

U.S. Military Entrance Processing Command, *Technical Supplement to the Counselor's Manual for the Armed Services Vocational Aptitude Battery Form-14,* North Chicago, IL, July 1985.

Waters, Brian K., Dana H. Lindsley, *Ability of Military Recruits: 1950 to 1994 and Beyond,* presentation at 15th Biennial Applied Behavioral Sciences Symposium-USAFA, Assistant Secretary of Defense (Force Management Policy), April 1996.